

CS5056 Data Analytics  
Homework Assignment 3  
Due: March 4, 2020

Exploratory Data Analysis to your dataset

You may answer HWA3 in teams of two or individually

Do an Exploratory Data Analysis and Visualization exercise over the dataset of your choice and use the RStudio platform to perform the following tasks:

- Select any one attribute to calculate its mean and standard deviation and visualize the results
- Select three variables, do a one-way analysis of variance and visualize the results
- Select two variables, do a Pearson correlation analysis, a Spearman correlation analysis, and visualize both
- Visualize the data set using one or various graph tools like
  - Scatter Plot
  - Histogram
  - Bar & Stack Bar Chart
  - Box Plot
  - Area Chart
  - Heat Map
  - Other graphing tools
- Propose a research question, hypothesis, or business problem you would like to investigate from the data understanding exercise obtained in analyzing and visualizing the data

Write a (partial) review article to explain your findings following the CRISP-DM methodology and explain in detail the Business understanding and data understanding phases of CRISP-DM. Again, use the IEEE template <https://www.overleaf.com/latex/templates/ieee-conference-template-example/nsncsyjfmpxy> and adhere to the following outline:

- Title of your review
- Autor/Authors (yourselves)
- Abstract
- Introduction

- Methodology
  1. Business understanding (Detailed)
  2. Data understanding and preparation (Detailed)
  3. Modelling (To do data understanding and preparation)
  4. Evaluation
  5. Preparation
- Discussion (About business and data understanding preparation)
- Conclusion (What did you learn from this exercise)
- References

Keep the length of the review article to no more than six pages.